EXHIBIT 3

SENSORY EVALUATION OF FOOD

PRINCIPLES AND PRACTICES

HARRY T. LAWLESS
CORNELL UNIVERSITY
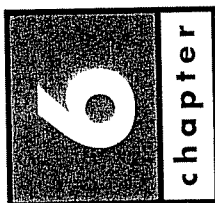
HILDEGARDE HEYMANN
UNIVERSITY OF MISSOURI

CHAPMAN & HALL

I⊕P® **International Thomson Publishing**
Thomson Science
New York • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • Pacific Grove • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

---

Food Science Texts Series

Series Editor
Dennis R. Heldman, University of Missouri

Editorial Board

Richard W. Hartel	University of Wisconsin
Hildegarde Heymann	University of Missouri
Joseph H. Hotchkiss	Cornell University
James M. Jay	University of Nevada Las Vegas
Kenneth Lee	Ohio State University
Steven J. Mulvaney	Cornell University
S. Suzanne Nielsen	Purdue University
Merle D. Pierson	Virginia Polytechnic Institute and State University
J. Antonio Torres	Oregon State University

Harry T. Lawless and Hildegarde Heymann
Sensory Evaluation of Food: Principles and Practices

FOOD SCIENCE TEXTS SERIES

Cameron Hackney, Merle D. Pierson, and George J. Banwart, *Basic Food Microbiology, 3rd Edition* (1998)

James M. Jay, *Modern Food Microbiology, 5th Edition* (1996)

Dennis R. Heldman and Richard W. Hartel, *Principles of Food Processing* (1997)

Norman G. Marriott, *Essentials of Food Sanitation* (1997)

S. Suzanne Nielsen, *Food Analysis, 2nd Edition* (1998)

Norman N. Potter and Joseph H. Hotchkiss, *Food Science, 5th Edition* (1995)

Romeo T. Toledo, *Fundamentals of Food Process Engineering, 3rd Edition* (1998)

Vickie A. Vaclavik, *Essentials of Food Science* (1997)

Ernest R. Vieira, *Elementary Food Science, Fourth Edition* (1996)

– – and O'Mahony, M. (1995). Probabilistic models for sequential taste effects in triadic choice. Journal of Experimental Psychology, Human Perception and Performance, 21 1088–1097.

Ferdinandus, A., Oosterom-Kleijngeld, I., and Runneboom, A.J.M. 1970. Taste testing," MBAA Technical Quarterly, 7(4) : 210–227.

Finney, D.J. 1971. Probit Analysis, 3d ed. Cambridge University Press, Cambridge.

Frijters, J.E.R. 1979. The paradox of the discriminatory nondiscriminators resolved. Chemical Senses, 4, 355–358.

– – –, Kooistra, A., and Vereijken, P.F.G. 1980. Tables of d' for the triangular method and the 3-AFC signal detection procedure. Perception and Psychophysics, 27(2), 176–178.

Green, D.M. and Swets, J.A. 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Lawless, H.T., and Schlegel, M.P. 1984. Direct and indirect scaling of taste—odor mixtures. Journal of Food Science, 49, 44–46.

Macmillan, N.A., and Creelman, C.D. 1991. Detection Theory: A User's Guide. Cambridge University Press, Cambridge.

McBurney, D.H. 1976. Signal detection theory and pain. Anesthesiology, 44, 356–358.

Meilgaard, M., Civille, G.V., and Carr, B.T. 1991. Sensory Evaluation Techniques, 2d ed. CRC, Boca Raton, FL.

O'Mahony, M.A. 1979. Short-cut signal detection measures for sensory analysis. Journal of Food Science, 44(1), 302–303.

– – – and Odbert, N. 1985. A comparison of sensory difference testing procedures: sequential sensitivity analysis and aspects of taste adaptation. Journal of Food Science, 50, 1055.

– – –, Masuoka, S., and Ishii, R. 1994. A theoretical note on difference tests: models, paradoxes and cognitive strategies. Journal of Sensory Studies, 9, 247–272.

Stillman, J.A., and Irwin, R.J. 1995. Advantages of the same-different method over the triangular method for the measurement of taste discrimination. Journal of Sensory Studies, 10, 261–272.

Thurstone, L.L. 1927. A law of comparative judgment. Psychological Review, 34, 273–286.

Ura, S. 1960. Pair, triangle and duo-trio test. Reports of Statistical Application Research, Japanese Union of Scientists and Engineers, 7, 107–119.

MEASUREMENT OF SENSORY THRESHOLDS

chapter 6

A light may be so weak as not sensibly to dispel the darkness, a sound so low as not to be heard, a contact so faint that we fail to notice it. In other words, a finite amount of the outward stimulus is required to produce any sensation of its presence at all. This is called by Fechner the law of the threshold— something must be stepped over before the object can gain entrance to the mind. —William James, 1913, p. 16

INTRODUCTION: THE APPEAL OF THRESHOLD CONCEPTS

One of the earliest characteristics of human sensory function to be measured was the absolute threshold. The absolute or detection threshold was seen as an energy level below which no sensation would be produced by a stimulus and above which a sensation would reach consciousness. The concept of threshold was central to Fechner's psychophysics. His integration of Weber's law to produce the first psychophysical relationship ($S = k \log I$) in fact depends formally on the physical intensity, I, being measured with the absolute threshold as the unit (Boring, 1946). Early physiologists like Weber and Fechner used the classical method of limits to measure this point of discontinuity or beginning of the psychophysical function. In the method of limits, the energy or physical intensity level would be raised and lowered, and the average point at which the observer changed response from "no sensation" to "yes, I perceive something" would be taken as the threshold. This specification of the minimum energy level required

the intensity judgments for individual differences in the range of the response scale that was used. Thus there is a moderate negative correlation of sensitivity and rated intensity when one examines the data across a highly variable group, as is the case with specific anosmia or tasting PTC bitterness. The correlation is negative since higher thresholds lead to lower sensitivity and lower rated intensity.

Group data from discrimination tests can also be used to estimate thresholds from the chance-corrected percent correct. Another approach using the ascending forced-choice to estimate a threshold is to work solely from group percent correct at each concentration level, rather than averaging thresholds of individuals. In this approach, each concentration level is treated like a simple discrimination test, e.g., a triangle procedure. In a series of tests, different concentrations are presented, so that at some concentration level a predetermined criterion for threshold is surpassed. It is critical in these tests to avoid the temptation to use statistical significance as a criterion for threshold. The statistical significance of any result depends not only on the percent but also on the number of judges participating. So using statistical significance as a criterion for threshold leads to the nonsensical result that lower thresholds are estimated as the number of testers increases (an irrelevant variable).

An alternative approach is to use some arbitrary level after adjustment for chance guessing, such as 50% above chance as the threshold (Brown et al., 1978; ASTM, 1991b). This makes use of Abbott's formula (Finney, 1971):

$$\text{Adjusted proportion correct} = \frac{observed\ proportion - chance}{1 - chance} \qquad (6.4)$$

In other words, the 50% above-chance level for the triangle test would be 66.7% correct in the experimental result and 75% correct for the duo–trio test or any paired test. This approach was taken, for example, by Antinone et al. (1994) in estimating the threshold for perception of diacetyl added to cottage cheese.

While this approach is superior to using statistical significance as a criterion for threshold, it is also not without liabilities, as discussed in Chapter 5. The primary difficulty is that different forms of discrimination tests will require the same adjusted proportion as long as they have the same chance probability. So, for example, the triangle test (pick the odd sample) will require 66.7% correct as will the 3-AFC test (pick the strongest sample). However, the oddity test is more difficult, and according to signal detection and Thurstonian scaling considerations requires a bigger sensory difference to get to this same level of performance (Ennis, 1995). The Thurstonian models, on the other hand, take this into account and can pro-

vide alternative measures of sensory difference that do not require an arbitrary cutpoint from Abbott's formula. Nonetheless, this is one practical approach to threshold measurement from group data, as long as the experimenter remembers that the obtained threshold is once again specific to that test method.

In summary, the ascending forced-choice method is a reasonably useful compromise between the need to precisely define a threshold level and the problems encountered in sensory adaptation and observer fatigue when extensive measurements are made. However, the user of an ascending forced-choice procedure should be aware of the many methodological variables that can affect the obtained threshold value. The following variables will affect the measured value: the number of alternatives (both targets and blanks), the stopping rule or the number of correct steps in a row required to establish a threshold, the number of replicated correct trials required at any one step, and the rule to determine at what level of concentration steps the threshold value is assigned. For example, the individual threshold might be assigned at the lowest level correct, the geometric mean between the lowest level correct and highest level incorrect. Other specific factors include the chosen step size of concentration units (factors of 2 or 3 are common in taste and smell), the method of averaging or combining replicated ascending runs on the same individual (geometric means are common but not universal), and finally the method of averaging or combining group data.

ALTERNATIVE APPROACHES: RATED DIFFERENCE, ADAPTIVE PROCEDURES, SIGNAL DETECTION

Rated Difference from Control

Another practical procedure for estimating threshold has involved the use of rated difference scales, where a sample containing the to-be-recognized stimulus is compared to some control or blank stimulus (Lundahl et al., 1986; Brown et al., 1978). In these procedures, ratings for the sensory difference from the control sample will increase as the intensity of the target gets stronger. Some point along the plotted ratings must be taken as threshold. In some variations on this method, a blind control sample is also rated. This provides the opportunity to estimate some baseline of scale usage or a kind of false-alarm estimate based on the ratings (often nonzero) of the control against itself. This comparison will often get nonzero difference estimates due to the moment-to-moment variability in sensations, akin to variation in the noise variability in signal detection experiments.